



# The importance of amino acid composition in natural AMPs: an evolutionary, structural, and functional perspective

**Biswajit Mishra and Guangshun Wang\***

Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE, USA

\*Correspondence: [gwang@unmc.edu](mailto:gwang@unmc.edu)

## Edited by:

Mark W. Robinson, University of Technology Sydney, Australia

## Reviewed by:

Andrew Tasman Hutchinson, University of Technology Sydney, Australia

Antimicrobial peptides (AMPs) are critical components of natural host defense systems against infectious pathogens (Zasloff, 2002; Boman, 2003; Hancock and Sahl, 2006). They are ubiquitous in nature and have been found in nearly all forms of life, ranging from single-celled bacteria to multicellular organisms such as plants and animals. AMPs are short peptides (5–100 amino acids) with an average net charge of +3 (Wang, 2010). They can display broad or narrow-spectrum antimicrobial activities. The fact that AMPs are effective against multidrug resistance pathogens, including suppression of biofilm formation, deserves our attention (Menousek et al., 2012). In addition to direct bacterial elimination, these peptides have regulatory effects on immune systems. Consequently, AMPs are also referred to as host defense peptides (Hancock and Sahl, 2006). To decode the key elements behind the functional diversity of AMPs, we have been taking time and efforts in constructing a comprehensive database that annotates such information. The first version of the Antimicrobial Peptide Database (APD; <http://aps.unmc.edu/AP/main.html>) was established in 2003 (Wang and Wang, 2004) and the database has since been further developed (Wang et al., 2009). The APD contained 1973 entries as of May 2012. To facilitate our bioinformatic analysis, we will register a peptide into the APD if it is (1) from natural sources; (2) with minimal inhibitory concentration (MIC) of less than 100  $\mu$ M or 100  $\mu$ g/mL; (3) less than 100 amino acid residues; and (4) with a characterized amino acid sequence (Wang, 2010). The APD allows users to extract important parameters (e.g., charge, hydrophobicity, motif, and structure) that determine peptide function. In particular, our database enables the generation of the amino acid composition for a select peptide

or a family of AMPs with a common feature. This bioinformatic tool thus uncovers the amino acid use in natural AMPs from different sources, with different functions, or three-dimensional structures. This opinion article highlights the critical roles of the amino acid composition in naturally occurring AMPs in terms of evolutionary, structural, and functional significance. Moreover, its application in designing and predicting new AMPs will also be discussed.

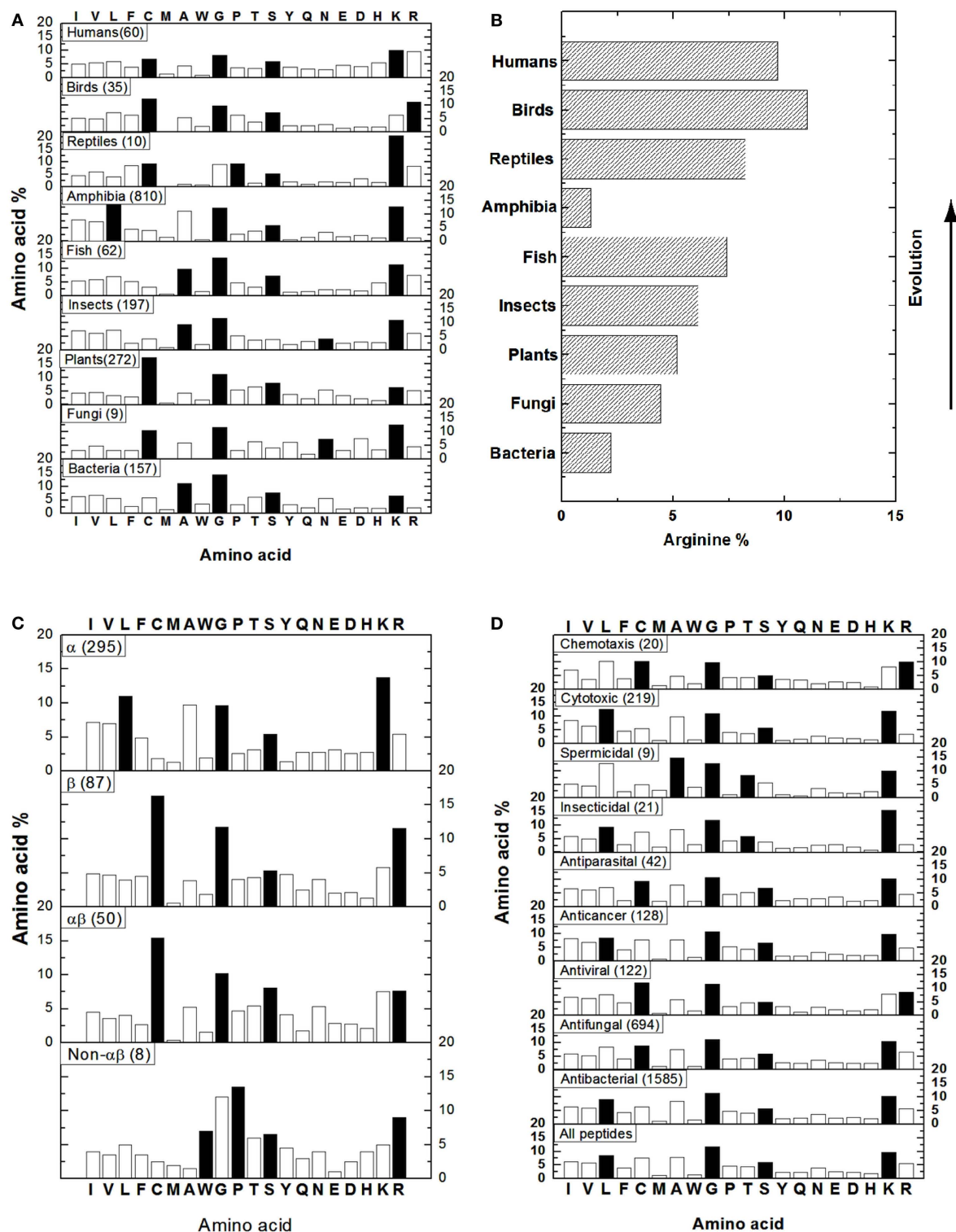
## EVOLUTIONARY PERSPECTIVE

To get an idea of AMPs in different kingdoms, we obtained the amino acid composition profiles of these peptides from bacteria, fungi, plants, insects, fish, amphibia, reptiles, birds, and humans (Figure 1A) by performing source search in the APD (Wang et al., 2009). In our database, the 20 standard amino acids are classified into four groups: hydrophobic (I, V, L, F, C, M, A, and W), GP (G and P), polar (T, S, Y, Q, and N), and charged (E, D, H, K, and R; Wang and Wang, 2004). In Figure 1, the dominant amino acids (highest percentages) in the four groups are represented as solid bars. For bacterial AMPs (i.e., bacteriocins), alanines (A) are the most preferred hydrophobic amino acid while residues G, S, and K are the most abundant in the other three groups (Figure 1A). Similarly, A is also a dominating hydrophobic residue in AMPs from insects or fish. In amphibian AMPs, L is the most abundant hydrophobic residue. In contrast, C is the major hydrophobic residue in AMPs from fungi, plants, and birds, probably due to the dominance of disulfide bonded defensin-like molecules. In human or reptile AMPs, C is comparable to other hydrophobic residues (e.g., L), probably reflecting the diversity in peptide sequences. For example, the known human AMPs are defensins, cathelicidins,

histatins, and  $\beta$ -amyloid peptides. Like the case of bacteria, G, S, and K are usually the dominant residues in the other three amino acid groups in Figure 1A. Exceptions are as follows. In the case of reptile AMPs, G and P are comparable. For AMPs from fungi and insects, the level of N is higher than or similar to S. Different from other kingdoms, birds select arginines (R) as the main charged amino acid, whereas arginines and lysines (K) are comparable in human AMPs. Based on the above description, it is clear that the dominant hydrophobic amino acids differ in various kingdoms, while residues G, S, and K are generally preferred amino acids in natural AMPs from nearly all the kingdoms (Figure 1A). The variations in the dominant amino acids in the hydrophobic group are an important observation and could suggest the preference of specific types of AMPs in certain kingdoms. In addition, one of the most important aspects is the observation made by Torrent et al. (2011) also on the basis of the APD (Wang et al., 2009). They found that higher organisms tend to incorporate R more frequently than K except amphibians (Figure 1B). The authors attributed this phenomenon to the possible emergence of the adaptive immune systems and the arginine-rich AMPs may well play an important role in modulating the immune system and in linking the innate and adaptive immune systems.

## STRUCTURAL INSIGHT

It is now clear that AMPs can adopt a variety of fascinating scaffolds, ranging from linear to circular. However, there are only four types of structures based on secondary structures:  $\alpha$ ,  $\beta$ ,  $\alpha\beta$ , and non- $\alpha\beta$  (Wang, 2010). The  $\alpha$  family consists of AMPs with  $\alpha$ -helical structures, while the  $\beta$  family comprises AMPs with a  $\beta$ -sheet structure. Another two families can be understood



**FIGURE 1 | Amino acid composition analysis of natural AMPs based on source (A,B), structure (C), and activity (D). (B) Shows the arginine percentages of AMPs from different life domains. In each case, the number of peptides included in the bioinformatic analysis is given in parentheses. The solid bars represent the most abundant amino**

**acids in the hydrophobic (I, V, L, F, C, M, A, W), GP (G and P), polar (T, S, Y, Q, N), and charged (E, D, H, K, R) groups. Data were obtained in May 2012 (total peptides: 1973) from <http://aps.unmc.edu/AP/main.php> by using the search and statistical analysis functions of the APD (Wang et al., 2009).**

accordingly:  $\alpha\beta = \alpha + \beta$  and non- $\alpha\beta = \text{no } \alpha \text{ and no } \beta$  structure. Representative structures for these four families can be viewed at the face page of the APD website above. The APD has also annotated those AMPs with determined 3D structures, which form the basis for our amino acid analysis. The peptides belonging to the  $\alpha$  family are widely distributed in bacteria and animals, while most of the plant AMPs, such as cyclotides and defensins, belong to the  $\beta$  family. The  $\alpha\beta$  members occur in all kingdoms, including bacteria, plants and animals, but the non- $\alpha\beta$  AMPs are much less frequent and are only confined to the animal kingdom at present. Depending on the structural family, the dominant amino acid in each amino acid group differs (**Figure 1C**). The  $\alpha$  helical AMPs prefer L as the major hydrophobic amino acid while K is selected as the charged amino acid. On contrary, the  $\beta$  stranded AMPs are dominated by C that determines the polypeptide fold. Meanwhile, it prefers R instead of K as the charged amino acid. Likewise, the  $\alpha\beta$  family has a high content of C as the hydrophobic component required for peptide folding. However, it possesses equal amounts of R and K. Finally, the non- $\alpha\beta$  family is generally composed of AMPs that are rich in particular amino acids such as tryptophan (W), proline (P), and R. The G and S are the other two preferred amino acids in all the families (**Figure 1C**). It is evident that amino acid composition is related to 3D structure of natural AMPs (Wang et al., 2009).

## FUNCTIONAL IMPLICATIONS

Natural AMPs with either narrow or broad-spectrum activity have been reported. Moreover, there is overlap in the activity spectrum of some AMPs (Zasloff, 2002; Hancock and Sahl, 2006). Such an activity spectrum for each AMP has been annotated in the APD. This includes antibacterial, antifungal, antiviral, antiparasitic, insecticidal, spermicidal, anticancer, cytotoxic (e.g., hemolytic), and chemotactic activity (Wang and Wang, 2004; Wang et al., 2009). **Figure 1D** shows the distribution of amino acids based on peptide activity. Except for spermicidal and insecticidal peptides, where T is preferred in the polar group, amino acids G and S are the two representative residues in the GP and polar amino acid groups

in all the cases. For hydrophobic residues, L is the dominant amino acid for AMPs with cytotoxic, insecticidal, anticancer, or antibacterial activity. There is a subtle difference between AMPs active against Gram-positive and Gram-negative strains only. AMPs active against Gram-positive bacteria have similar contents of C and L, whereas those against only Gram-negative AMPs have higher L and lower C contents (not shown). Residue C is dominant in the hydrophobic group for AMPs with chemotactic, antiparasitic, antiviral, or antifungal activity, suggesting the existence of a significant number of disulfide bonded molecules. In the case of spermicidal AMPs, residue A is the major hydrophobic amino acid. Finally, while lysines are usually the positively charged residue, arginines are clearly important in AMPs with chemotactic and antiviral activities. Therefore, the amino acid composition plays a role in determining peptide activity as well. For example, anticancer peptides are rich in L, G, S, and K, whereas chemotactic peptides have high C, G, T, and R contents.

## APPLICATIONS OF ABUNDANT AMINO ACIDS IN PEPTIDE DESIGN

As recently summarized by Wang (2010), there are various methods in designing new AMPs, ranging from template optimization, motif hybridization, sequence shuffling/library screening, to rationale design. It has been recognized that parameters such as charge and hydrophobicity play a tremendous role in determining AMP activity (Zasloff, 2002; Hancock and Sahl, 2006; Wang, 2010). Our opinion is that the abundant residues identified in amino acid composition profiles of AMPs (**Figure 1**) can be used to design a specific peptide with the *desired* activity. Indeed, we succeeded in designing a 19-residue peptide using only residues G, L, and K. GLK-19 is active against *E. coli* but not *S. aureus* (Wang et al., 2009) or HIV-1. Since antiviral AMPs prefer arginines (**Figure 1D**), we obtained an anti-HIV peptide GLR-19 after the conversion of lysines in GLK-19 to arginines (Wang et al., 2010). In addition, because C is the dominant hydrophobic residue in antiviral peptides, we further improved anti-HIV activity of the peptide after introduction of a pair of cysteines to GLR-19 between residues 4 and 16 (Wang

et al., 2011). Therefore, we succeeded in modulating peptide activity by varying the amino acid composition. We also propose that the prediction interface of the APD can be improved based on the abundant amino acids identified herein.

## A SUMMARY OF OPINIONS

The construction of the APD made it possible for us to extract the amino acid composition information in natural AMPs for the first time. Further classification of AMPs and the update of our database made the extracted parameters more informative (Wang et al., 2009). We propose that the amino acid composition plays an important role in terms of evolution, structure, and function of natural AMPs. The overall picture for natural AMPs is shaped through evolution. For example, the preference of arginines in the AMPs in higher organisms (**Figure 1B**) is proposed to be significant in the emergence of adaptive immune systems (Torrent et al., 2011) and probably also confers the regulatory and integrative role to natural AMPs in host defense. We demonstrated that arginines are more effective in targeting MRSA or HIV-1 (Wang et al., 2010, 2012). The amino acid composition appears to directly determine the various structural scaffolds of natural AMPs (**Figure 1C**). In the case of amphibians, the dominance of G, L, A, and K determines a helical conformation, leading to a natural recombinant library of peptides (up to ~100 in each frog) achieved by presenting varying other amino acids on the same helical structure backbone (Wang et al., 2009). Likewise, plant cyclotides are rich in C, G, T/S, and K that determines a universal  $\beta$ -sheet containing scaffold (Wang, 2010). Again, nature has created a natural recombinant library of cyclotides by introducing other amino acids to various loop regions. The same strategy is now utilized to generate new cyclotides with a desired biological function via segment grafting (Craik et al., 2012). Amino acid compositions may determine the mechanisms of action of natural AMPs. As we noticed previously, plant cyclotides and bacterial lantibiotics have different structural folds but similar amino acid composition profiles (Wang, 2010). Interestingly, certain lantibiotics can bind phosphatidylethanolamines (PE; Zhao, 2011; Ökseli

et al., 2011), so can cyclotides (Henriques et al., 2011). There is now converged view regarding the mechanism of action of proline-rich AMPs. They can cross bacterial membranes and associate with heat-shock proteins (Scocchi et al., 2011). The abundant amino acids elucidated from the APD (Figure 1) are helpful for both prediction and design of new AMPs (Wang, 2010). In our opinion, database-guided design is preferred over library screening due to its cost effectiveness by synthesizing only few peptides. Future database annotations and expansion are anticipated to further improve the accuracy of the amino acid composition profiles, thereby opening the door to other potential applications.

## ACKNOWLEDGMENT

We appreciate NIAID/NIH funding (R56AI081975 and R21AI082689) to Guangshun Wang during this study.

## REFERENCES

- Boman, H. G. (2003). Antibacterial peptides: basic facts and emerging concepts. *J. Intern. Med.* 254, 197–215.
- Craik, D. J., Swedberg, J. E., Mylne, J. S., and Cemazar, M. (2012). Cyclotides as a basis for drug design. *Expert Opin. Drug Discov.* 7, 179–194.
- Hancock, R. E. W., and Sahl, H. (2006). Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat. Biotechnol.* 24, 1551–1557.
- Henriques, S. T., Huang, Y. H., Rosengren, K. J., Franquelim, H. G., Carvalho, F. A., Johnson, A., Souza, S., Tachedjian, G., Castanho, M. A., Daly, N. L., and Craik, D. J. (2011). Decoding the membrane activity of the cyclotide kalata B1: the importance of phosphatidylethanolamine phospholipids and lipid organization on hemolytic and anti-HIV activities. *J. Biol. Chem.* 286, 24231–24241.
- Menousek, J., Mishra, B., Hanke, M. L., Heim, C. E., Kielian, T., and Wang, G. (2012). Database screening and in vivo efficacy of antimicrobial peptides against methicillin-resistant *Staphylococcus aureus* USA300. *Int. J. Antimicrob. Agents* 39, 402–406.
- Ökesli, A., Cooper, L. E., Fogle, E. J., and van der Donk, W. A. (2011). Nine post-translational modifications during the biosynthesis of cinnamycin. *J. Am. Chem. Soc.* 133, 13753–13760.
- Scocchi, M., Tossi, A., and Gennaro, R. (2011). Proline-rich antimicrobial peptides: converging to a non-lytic mechanism of action. *Cell. Mol. Life Sci.* 68, 2317–2330.
- Torrent, M., Valle, J., Nogus, M. V., Boix, E., and Andreu, D. (2011). The generation of antimicrobial peptide activity: a trade-off between charge and aggregation? *Angew. Chem. Int. Ed.* 50, 10686–10689.
- Wang, G. (2010). *Antimicrobial Peptides: Discovery, Design and Novel Therapeutic Strategies*. Wallingford: CABI.
- Wang, G., Buckheit, K. W., Mishra, B., Lushnikova, T., and Buckheit, R. W. (2011). De novo design of antiviral and antibacterial peptides with varying loop structures. *J. AIDS Clin. Res.* S2, 003.
- Wang, G., Epand, R. F., Mishra, B., Lushnikova, T., Thomas, V. C., Bayles, K. W., and Epand, R. M. (2012). Decoding the functional roles of cationic side chains of the major antimicrobial region of human cathelicidin LL-37. *Antimicrob. Agents Chemother.* 56, 845–856.
- Wang, G., Li, X., and Wang, Z. (2009). APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res.* 37, D933–D937.
- Wang, G., Watson, K. M., Peterkofsky, A., and Buckheit, R. W. Jr. (2010). Identification of novel human immunodeficiency virus type 1 inhibitory peptides based on the antimicrobial peptide database. *Antimicrob. Agents Chemother.* 54, 1343–1346.
- Wang, Z., and Wang, G. (2004). APD: the antimicrobial peptide database. *Nucleic Acid Res.* 32, D590–D592.
- Zasloff, M. (2002). Antimicrobial peptides of multicellular organisms. *Nature* 415, 389–395.
- Zhao, M. (2011). Lantibiotics as probes for phosphatidylethanolamine. *Amino Acids* 41, 1071–1079.

Received: 25 June 2012; accepted: 9 July 2012; published online: 31 July 2012.

Citation: Mishra B and Wang G (2012) The importance of amino acid composition in natural AMPs: an evolutionary, structural, and functional perspective. *Front. Immun.* 3: 221. doi: 10.3389/fimmu.2012.00221

This article was submitted to *Frontiers in Molecular Innate Immunity*, a specialty of *Frontiers in Immunology*.

Copyright © 2012 Mishra and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution, and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.